UNIVERSITÀ DEGLI STUDI DI TORINO

Global gene flow releases invasive plants from environmental constraints on genetic diversity

Annabel L. Smith[1,2,*], Trevor R. Hodkinson[3], Jesus Villellas[4], Jane A. Catford[5], Anna Mária Csergő[1,6], Simone P. Blomberg[7], Elizabeth E. Crone[8], Johan Ehrlén[9], Maria B. Garcia[10], Anna-Liisa Laine[11,12], Deborah A. Roach[13], Roberto Salguero-Gómez[14], Glenda M. Wardle[15], Dylan Z. Childs[16], Bret D. Elderd[17], Alain Finn[1], Sergi Munné-Bosch[18,19], Maude E. A. Baudraz[1], Judit Bódis[20], Francis Q. Brearley[21], Anna Bucharova[22,23], Christina M. Caruso[24], Richard P. Duncan[25], John M. Dwyer[26,27], Ben Gooden[28,29], Ronny Groenteman[30], Liv Norunn Hamre[31], Aveliina Helm[32], Ruth Kelly[1], Lauri Laanisto[33], Michele Lonati[34], Joslin L. Moore[35], Melanie Morales[18,36], Siri Lie Olsen[37], Meelis Pärtel[32], William K. Petry[38], Satu Ramula[39], Pil U. Rasmussen[9,40], Simone Ravetto Enri[34], Anna Roeder[41,42], Christiane Roscher[41,42], Marjo Saastamoinen[43,44], Ayco J. M. Tack[9], Joachim Paul Töpper[45], Gregory E. Vose[46], Elizabeth M. Wandrag[25,47], Astrid Wingler[48] and Yvonne M. Buckley[1]

[1]Zoology, School of Natural Sciences, Trinity College Dublin, Dublin 2, Ireland. [2]School of Agriculture and Food Science, University of Queensland, Gatton 4343, Australia. [3]Botany, School of Natural Sciences, Trinity College Dublin, Dublin 2, Ireland. [4]Departamento Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales, MNCN-CSIC, E-28006 Madrid, Spain. [5]Department of Geography, King's College London, 30 Aldwych, London, WC2B 4BG, UK. [6]Department of Botany and Soroksár Botanical Garden, Faculty of Horticultural Science, Szent István University, 1118 Budapest, Villány Street 29-43 Hungary. [7]School of Biological Sciences, University of Queensland, Brisbane, Queensland 4072, Australia. [8]Department of Biology, Tufts University, Medford MA 02145 USA. [9]Department of Ecology, Environment and Plant Sciences, Stockholm University, SE-106 91 Stockholm, Sweden. [10]Pyrenean Institute of Ecology (CSIC), Avda. Montañana, 1005, Zaragoza 50059, Spain. [11]Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland. [12]Research Centre for Ecological Change, Faculty of Biological and Environmental Sciences, PO Box 65 (Viikinkaari 1), FI-00014 University of Helsinki, Finland. [13]Department of Biology, University of Virginia, Charlottesville, Virginia, 22904, USA. [14]Department of Zoology, University of Oxford, 11a Manfield Rd, OX1 3SZ, Oxford, UK. [15]School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia. [16]Department of Animal and Plant Sciences, University of Sheffield, Sheffield, S10 2TN, UK. [17]Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana 70803, USA. [18]Department of Evolutionary Biology, Ecology and Environmental Sciences, University of Barcelona, Barcelona, Spain. [19]Institut de Recerca de la Biodiversitat (IrBio), University of Barcelona, Barcelona, Spain. [20]Georgikon Faculty, University of Pannonia, Deák F. utca 16. H-8360, Hungary. [21]Department of Natural Sciences, Manchester Metropolitan University, Chester Street, Manchester, M1 5GD, UK. [22]Plant Evolutionary Ecology, Institute of Evolution and Ecology, University of Tübingen, Germany. [23]Ecosystem and Biodiversity Research Group, Institute of Landscape Ecology, University of Münster, Germany. [24]Department of Integrative Biology, University of Guelph, Guelph, Ontario N1G 2W1 Canada. [25]Institute for Applied Ecology, University of Canberra, ACT 2617, Australia. [26]The University of Queensland, St Lucia, QLD, Australia 4072. [27]CSIRO Land & Water, EcoSciences Precinct, Dutton Park, QLD, Australia 4102. [28]CSIRO Health and Biosecurity, Black Mountain, ACT 2601, Australia. [29]School of Earth, Atmospheric and Life Sciences, Faculty of Science, Medicine and Health, University of Wollongong, Wollongong, Australia. [30]Manaaki Whenua - Landcare Research, 54 Gerald Street, Lincoln, 7608, New Zealand. [31]Department of Environmental Sciences, Western Norway University of Applied Sciences, Røyrgata 6, N-6856 Sogndal, Norway. [32]Institute of Ecology and Earth Sciences, University of Tartu, Lai 40, 51005 Tartu, Estonia. [33]Chair of Biodiversity and Nature Tourism, Estonian University of Life Sciences, Kreutzwaldi 5, 51006, Tartu, Estonia. [34]Department of Agricultural, Forest and Food Science, University of Torino, Largo Braccini 2, 10015 Grugliasco (TO) Italy. [35]School of Biological Sciences, Monash University, Clayton Campus VIC 3800, Australia. [36]Research Group of Plant Biology under Mediterranean Conditions, Faculty of Biology, University of Balearic Islands, cta. Valldemossa km 7,5, 07122, Palma de Mallorca, Spain. [37]Norwegian Institute for Nature Research, Gaustadalléen 21, N-0349 Oslo, Norway. [38]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544 USA. [39]Department of Biology, University of Turku, 20014 Turku, Finland. [40]The National Research Centre for the Working Environment, Lersø Parkallé 105, 2100 København Ø, Denmark. [41]UFZ, Helmholtz Centre for Environmental Research, Physiological Diversity, Permoserstrasse 15, 04103 Leipzig, Germany. [42]German Centre for Integrative Biodiversity Research Halle-Jena-Leipzig (iDiv), Deutscher Platz 5e, 04318 Leipzig, Germany. [43]Helsinki Institute of Life Science, University of Helsinki, Finland. [44]Organismal and Evolutionary Research Programme, University of Helsinki, 00014 Helsinki, Finland. [45]Norwegian Institute for Nature Research, Thormøhlensgate 55, N-5006 Bergen, Norway. [46]Department of Ecology and Evolutionary Biology, University of California, Irvine, Irvine, CA 92697, USA. [47]University of New England, Armidale, NSW 2351, Australia. [48]School of Biological, Earth & Environmental Sciences and Environmental Research Institute, University College Cork, Distillery Fields, North Mall, Cork, Ireland

*** Correspondence**: Annabel Smith

Email: annabel.smith@uq.edu.au

Current address: School of Agriculture and Food Science, University of Queensland, Gatton 4343, Australia

Phone: +617 5460 1692

ORCID IDs:

Annabel L. Smith: 0000-0002-1201-8713
Jesus Villellas: 0000-0001-7805-5683
Jane A. Catford: 0000-0003-0582-5960
Anna Mária Csergő: 0000-0003-3325-2995
Deborah A. Roach: 0000-0002-5273-5370
Roberto Salguero-Gómez: 0000-0002-6085-4433
Bret D. Elderd: 0000-0001-5853-1136
Alain Finn: 0000-0002-2209-9553
Sergi Munné-Bosch: 0000-0001-6523-6848
Maude E.A. Baudraz: 0000-0001-7252-8419
Christina M. Caruso: 0000-0001-7069-9572
John M. Dwyer: 0000-0001-7389-5528
Ben Gooden: 0000-0003-0575-9078
Liv Norunn Hamre: 0000-0002-9033-4260
Ruth Kelly: 0000-0001-7982-5993
Lauri Laanisto: 0000-0003-2215-7298
Michele Lonati: 0000-0001-8886-0328
Siri Lie Olsen: 0000-0002-4443-8261
Meelis Pärtel: 0000-0002-5874-0138
William K. Petry: 0000-0002-5230-5987
Satu Ramula: 0000-0001-7795-0352
Pil U. Rasmussen: 0000-0003-0607-4230
Simone Ravetto Enri: 0000-0002-3584-8031
Ayco J. M. Tack: 0000-0002-3550-1070
Joachim Paul Töpper: 0000-0002-6996-7223
Astrid Wingler: 0000-0003-4229-2497

Classification

BIOLOGICAL SCIENCES: Ecology

Keywords

Plant invasion | Adaptation | Global change | Population genetics | Demography

Author Contributions

YMB coordinated the PLANTPOPNET network. The founding steering committee (YMB, SPB, EEC, AMC, JE, MBG, A-LL, DAR, RS-G and GW) designed the PLANTPOPNET network and wrote the demographic census protocol, while the current steering committee (including DZC, BDE, AF, SM-B and JV) oversee network operation. ALS, YMB and TRH designed the concept, DNA data collection and analytical approach for the current study. ALS conducted all analyses and wrote the code for SNP filtering and analysis. ALS wrote the first manuscript draft with major contributions from YMB, TRH, JV, JAC and AMC. All authors (except SPB and JE) contributed data used in the current study. All authors made contributions to the final manuscript.

This PDF file includes:

Main Text
Figures 1 to 5
Table 1

Abstract

When plants establish outside their native range, their ability to adapt to the new environment is influenced by both demography and dispersal. However, the relative importance of these two factors is poorly understood. To quantify the influence of demography and dispersal on patterns of genetic diversity underlying adaptation, we used data from a globally-distributed demographic research network, comprising 35 native and 18 non-native populations of *Plantago lanceolata*. Species-specific simulation experiments showed that dispersal would dilute demographic influences on genetic diversity at local scales. Populations in the native European range had strong spatial genetic structure associated with geographic distance and precipitation seasonality. In contrast, non-native populations had weaker spatial genetic structure that was not associated with environmental gradients, but with higher within-population genetic diversity. Our findings show that dispersal caused by repeated, long-distance, human-mediated introductions have allowed invasive plant populations to overcome environmental constraints on genetic diversity, even without strong demographic changes. The impact of invasive plants may therefore increase with repeated introductions, highlighting the need to constrain future introductions of species even if they already exist in an area.

Significance Statement

We found that long-distance dispersal and repeated introductions by humans have shaped adaptive potential in a globally distributed invasive species. Some plant species therefore do not need strong demographic changes to overcome environmental constraints that exist in the native range; simply mixing genetic stock from multiple populations can provide an adaptive advantage. This work highlights the value of preventing future introduction events for problematic invasive species, even if the species already exists in an area.

Main Text

Introduction

Patterns of genetic diversity across a species' range arise from a complex interplay between the diversifying effect of demographic variation across landscapes with different selection pressures, and the homogenising effects of dispersal[1-3]. On one hand, variability in demographic performance influences genetic diversity through its influence on effective population size[4]. Short-lived, highly fecund species generally have higher levels of genetic diversity compared to species that are long-lived or have low fecundity[5,6]. On the other hand, dispersal modulates these relationships by facilitating gene flow between populations[7]. Gene flow from seed and pollen can increase genetic diversity and reduce genetic differences among populations. While the importance of these forces is widely accepted[8], there is uncertainty about the relative strength of demography and dispersal in shaping genetic structure across global environmental gradients[9,10].

For invasive species, the situation is even more complex because humans disrupt many of the natural processes that determine genetic diversity (Fig. 1). For example, repeated introductions and long-distance dispersal by humans can release invasive plant species from demographic constraints, such as those imposed by the colonisation-competition tradeoff[11]. Invasive species might also overcome climatic constraints on phenotypic traits as a result of rapid adaptation to new environments[12] or non-adaptive processes such as repeated introductions which can swamp locally adapted phenotypes[13]. Thus, emerging evidence suggests that plants in their non-native range can break ecological 'rules' because they are not always constrained by the same biological and climatic forces that operate in their native range.

Some populations of invasive species lose genetic diversity during invasion through founder effects[14], but many have higher genetic diversity outside their native range[15,16]. The mechanisms

underlying this phenomenon include admixture (i.e. new genotypes arising from interbreeding among divergent source populations)[17], hybridisation[18], rapid mutation[19] and exposure of cryptic genetic variation[20]. Such increases in genetic diversity can enhance colonisation success[21] and adaptive potential[22] in invasive species. Demographic changes can also improve invasive plant performance[23], which is sometimes associated with release from natural enemies[24]. Unfortunately, demographic and genetic aspects of invasion are often analysed in isolation[25], in part because labour-intensive demographic studies are typically done at one or a few sites making them severely limited in spatial replication[26]. This means we lack understanding about the relative importance of demographic change and global dispersal on biological invasions[27,28].

Here, we present a demographically-informed analysis of neutral and putatively adaptive genetic diversity in *Plantago lanceolata* L. (Plantaginaceae), a common forb native to Europe and western Asia, which now has a cosmopolitan distribution (Fig. 2). *Plantago lanceolata* established in its non-native range through long-distance dispersal by humans[29], repeated introductions[30] and cultivation[31] – all processes that can increase genetic diversity and invasion success[15]. The overarching aim of the study was to analyse the influences of local demography and global dispersal patterns on genetic diversity in *P. lanceolata* and determine which of these pathways drives adaptive capacity. This knowledge is necessary to understand how future introduction events will influence the spread of invasive plants. This work was made possible by a globally-distributed demographic research network (PLANTPOPNET) and is, to our knowledge, the first analysis of genetic diversity at a global scale that integrates field-collected demographic data.

In addition to demographic data, we sampled DNA from 491 individuals including outgroups, cultivar lines and 53 naturally occurring populations across the native European range (n = 35) and the non-native range (n = 18) in southern Africa, Australasia and North America (Fig. 2). To address our main aim, three hypotheses were tested:

(H1) In absence of dispersal, increases in survival and fecundity will drive increases in genetic diversity. These effects will be diluted by dispersal between populations.

(H2) Patterns of spatial genetic structure among native populations will reflect dispersal limitations across environmental gradients. In the non-native range, gene flow arising from multiple introductions will disrupt spatial genetic structure observed in the native range.

(H3) Environmental influences on within-population genetic diversity will be explained by demographic variation (density, fecundity and empirical population growth rate). Repeated introductions into the non-native range and long-distance dispersal by humans will weaken this relationship (Fig. 1).

A genotypic simulation model, parameterised with empirical demographic data from *P. lanceolata*, was used to test H1. We then coupled field-collected demographic data (density, empirical population growth rate and fecundity) with single nucleotide polymorphism data (18,166 neutral and 3,024 putatively adaptive SNPs) to test H2 and H3.

Results and discussion

Hypothesis 1: Dispersal between populations will dilute demographic effects on genetic diversity

In two simulated populations unconnected by dispersal, with different rates of juvenile survival (o_j = 0.1 and 0.2) and female fecundity (seeds per plant, δ_F 1–100), higher juvenile survival led to greater genetic diversity (Fig. 3a). Above the threshold at which populations went extinct (δ_F = 15), genetic diversity increased sharply until δ_F was approximately 25. Above this point there was little influence of fecundity on genetic diversity (Fig. 3a). Population size at the end of the simulation was larger with higher juvenile survival (Fig. 3b). Thus, variation in female fecundity

appears to have less influence than juvenile survival in determining genetic diversity in *P. lanceolata*. When the two populations were connected by dispersal, differences in heterozygosity persisted until the number of migrants per generation exceeded 50,000 (Fig. 3c, d). This number is realistic in natural populations since reproductive individuals typically produce a minimum 20-100 seeds and migration refers to propagules dispersed before the recruitment process. Male fecundity was kept constant in the model as it is very high in *P. lanceolata* (10,000–54,000 pollen grains per anther[32]) and had no influence on genetic diversity.

The simulation result supports our prediction (H1) that demography would influence genetic diversity in *P. lanceolata* when dispersal barriers are present and that dispersal would dilute these effects. The simulation also suggests that juvenile survival is an important parameter controlling heterozygosity. When dispersal barriers are removed however, gene flow from pollen and seed will swamp local effects of juvenile survival on heterozygosity. We could therefore expect demographic effects on genetic diversity to become undetectable at the upper range of pollen and seed movement that occurs in *P. lanceolata*.

The increases in genetic diversity with juvenile survival (Fig. 3) might not confer an adaptive advantage since they reflect genetic diversity arising from neutral demographic processes. The relevance of this result however, is that there is enough demographic variability in *P. lanceolata* to shape neutral genetic structure, an assumption underlying the hypotheses in the rest of the study. Thus, we can expect juvenile survival to be the dominant demographic parameter underlying differences in *P. lanceolata* genetic diversity when dispersal is limited at local scales. At continental scales, genetic diversity is probably influenced less by juvenile survival when gene flow is high. This might be especially true in the non-native range where there has been a shorter history of local adaptation[33] and multiple human-mediated introductions (the human activity pathway, Fig. 1).

Hypothesis 2: Global gene flow from multiple introductions will disrupt spatial genetic structure

Admixture analysis of *P. lanceolata* genotypes with fastSTRUCTURE[34] revealed strong genetic structure in the native range and a high degree of admixture in the non-native range. The number of genetic clusters at Hardy-Weinberg Equilibrium (*K*) was between *K* = 6 (model complexity maximising marginal likelihood) and *K* = 13 (model components used to explain structure in the data). When *K* = 6, cultivar lines and outgroups (*P. coronopus* and *P. major*) formed two distinct clusters and the remaining four clusters were present in the native European range with clear spatial structure (Fig. 2). Greece, Italy, the Islands of the North Atlantic and Finland comprised almost ‘pure’ lines of these four clusters, while other European populations were admixed.

Genotypes of most non-native populations were admixed and there was relatively little spatial structure at a global scale (Fig. 2). This was supported by a significantly higher Diversity Score in the non-native range (model estimate, SE = 0.34, 0.04), compared to the native range (0.22, 0.03) (see SI Appendix, Fig. S6, *P* = 0.033). Italy and central France were the most similar source material for the dominant genotype in the non-native populations. Some cultivar stock was identified in the Spanish populations, possibly reflecting the Iberian source of material used to breed cultivars. The cultivars were developed in New Zealand, thus the presence of cultivar stock in that population might indicate mixing between the naturalised population and pasture plants (Fig. 2). At the upper range of *K*, further spatial structure was identified in Europe (e.g. at *K* = 13 Norway was differentiated from Finland), while the non-native populations still showed admixture of multiple, mostly Mediterranean sources (see SI Appendix, Fig. S1). The lack of spatial structure at a global scale was supported by Analysis of Molecular Variance (AMOVA) showing that genetic variation between the native and non-native range was only 2.2%, among individuals within populations was 10.7% and among populations within ranges was 11.4%. The remaining genetic variation (75.5%) accounted for individual heterozygosity.

The minimum number of colonising propagules required to produce the observed level of genetic diversity in non-native regions ($Prop_{min}$) depended on sample size (r = 0.99) and ranged from 5.35 in New Zealand to 49.95 in North America (Fig. 2). Multiple introductions were therefore required to produce observed levels of genetic diversity in the non-native ranges. Relative to sample size, $Prop_{min}$ ranged from 0.55 to 0.90 indicating that, in each region, more than half the sampled population was required to represent non-native genetic diversity. $Prop_{min}$ was based on the alleles present in the native range, but there were also a number of non-European alleles in each non-native region (12–159, Fig. 2). Thus, we either failed to sample the full extent of the source population (despite extensive sampling across Europe), or new genotypes were produced after colonisation. The latter explanation can arise through transgressive segregation[35] and is one mechanism by which invasive species adapt quickly to new environments. However, we also detected private alleles within sites in Europe (see SI Appendix, Table S1) so our sample does not represent the full range of genetic diversity in the species.

Genetic structure measured by F_{ST} (genetic differentiation between all pairs of populations) was stronger among populations in the native range (mean F_{ST} = 0.16) than the non-native range (mean F_{ST} = 0.09). To analyse the influence of environmental gradients on F_{ST}, we used three separate generalised dissimilarity models, one for each range type: native range, non-native range and the global population (native and non-native combined). The deviance explained by the native model was 74.3% (bootstrap CI = 68.6, 78.3) and two out of six variables fitted in the model had a significant influence on F_{ST} (Fig. 4, see SI Appendix, Fig. S2). Genetic distance increased with geographic distance (Fig. 4a) and sites with similar levels of precipitation seasonality were more genetically similar (Fig. 4b) after accounting for other variables in the model (see SI Appendix, Fig. S2). No variable significantly affected F_{ST} in the non-native range (deviance explained = 23.1%, bootstrap CI = 9.4, 34.1) or the global population (deviance explained = 10.9%, bootstrap CI = 7.25, 14.33) (see SI Appendix, Fig. S2). Geographic distance was included in each model to account for differences in spatial scale. Thus, if environmental influences on gene flow had persisted in the non-native range, they should have been detectable. Combined with the admixture analysis, these results support our prediction (H2) that multiple introductions from diverse source populations and long-distance dispersal can weaken environment–genetic structure relationships. *Plantago lanceolata* reproduces clonally as well as sexually and this flexible reproductive mode, combined with high admixture in the non-native range, suggests fast expansion after colonisation. This might allow the species to overcome ecological constraints, without the need for local adaptation[36].

In the native range of *P. lanceolata*, the increase in genetic distance with precipitation seasonality might partially reflect a historic biogeographical pattern (precipitation seasonality was correlated with longitude, r = 0.47). Historical processes occurring along both east-west and north-south axes shape contemporary genetic patterns in European plants. For example, glacial refugia in Iberia, Italy and the Balkans, were reflected in highly divergent lines of *Arabidopsis thaliana* south of the alpine barrier[37]. In our dataset, the Italian population was genetically distinct, while two eastern sites in Romania were highly differentiated and genetically related to Greece (Fig. 2). François et al.[37] also found evidence for an eastern refuge in *A. thaliana*. Further sampling into the continental Asian range of *P. lanceolata* would help uncover whether the observed patterns arose from movement with agriculture westward across Europe[38,39] or postglacial colonisers from the Balkans[40].

Hypothesis 3: Global gene flow will weaken demographic effects on genetic diversity within populations

We compared a series of linear models, including additive and interactive effects of range (native/non-native) to address the hypothesis that environmental influences on within population genetic diversity would differ between the native and non-native ranges (Dataset S1). Our results offered partial support for Hypothesis 3 because environmental gradients (characterised by mean temperature, temperature seasonality and mean precipitation) affected population growth rate,

fecundity and neutral and adaptive genetic diversity in native and non-native ranges of *P. lanceolata* (Fig. 5, see SI Appendix, Fig. S3). Our expectation, however, that genetic responses to the environment could be explained by demographic variation had little support (see SI Appendix, Fig. S3). Demographic variables responded to environmental gradients, but did not induce a response on genetic diversity when used as predictor variables. Demographic and genetic parameters within populations were best explained by environmental gradients and, in some cases, there were differences in the responses between native and non-native ranges.

The top-ranked models for population growth rate (Fig. 5a) and fecundity (Fig. 5b) had additive effects of mean temperature, responding similarly in the native and non-native ranges. Globally, warmer sites tended to have lower population growth rates and higher fecundity. Increases in fecundity can occur to offset lower survival in stressful environments[41], a phenomenon which has been recorded in other studies of *Plantago*[42,43]. There was also an additive effect of temperature seasonality on neutral genetic diversity (Fig. 5c), with highly seasonal sites having greater genetic diversity in the native and non-native ranges. Mean temperature and temperature seasonality were correlated (r = -0.36, p = 0.02, see SI Appendix, Fig. S4). Thus, the observed responses are best thought of as responses to an environmental gradient, with demographic and genetic parameters responding to different aspects of the gradient. High genetic diversity in highly seasonal sites might have been driven by increased fecundity, since we found some evidence of a positive relationship between fecundity and genetic diversity (see SI Appendix, Fig. S3g, Dataset S1).

Three of the top-ranked models included an interaction between environment and range, showing environmental effects in the native range but not the non-native range. Both neutral (Fig. 5d, bootstrap CI = 0.001, 0.010) and adaptive (Fig. 5f, bootstrap CI = 0.004, 0.021) genetic diversity decreased across a mean precipitation gradient in the native range, but not in the non-native range. Adaptive genetic diversity increased with temperature seasonality, but only in the native range (Fig. 5e, bootstrap CI = -0.021, -0.005). There was also support (ΔAICc < 2) for non-native populations having a weaker response to environmental gradients in terms of fecundity (see SI Appendix, Fig. S3a, b), population growth rate (see SI Appendix, Fig. S3c) and neutral genetic diversity (see SI Appendix, Fig. S3d). Taken together, these results suggest that non-native populations are not constrained by the same environmental forces as their native counterparts.

Population growth rate and neutral and adaptive genetic diversity were all higher in the non-native range (Fig. 5, Dataset S1), suggesting that invasive populations have a greater capacity for colonisation and adaptation. Higher population growth rates in non-native populations were probably driven by increases in survival rather than fecundity, since fecundity was lower in the non-native range (Fig. 5b, Dataset S1). Thus, our simulation experiments and our field data indicated stronger effects of survival than of fecundity on genetic diversity and population growth, respectively.

Increases in genetic diversity can arise when environmental heterogeneity drives population turnover through increases in sexual reproduction, population growth and survival[6,44]. In our study however, population growth was affected by mean temperature, not variability in temperature; cooler sites generally had higher rates of population growth across the first two demographic censuses. This is consistent with previous work showing that high mean temperature was associated with mortality in *P. lanceolata*[42]. Thus, we did not find a clear demographic explanation for the effect of temperature seasonality on genetic diversity. Temperature stability might have promoted clonality in *P. lanceolata*, leading to lower genetic diversity[45]. However, rates of sexual and clonal reproduction within species are often inversely related[46] and genetic diversity was unaffected by rates of sexual reproduction in our study. The influence of global variation in clonality on genetic diversity needs further investigation, particularly because clonality combined with sexual reproduction can increase invasion success[36].

Our prediction that environmental effects on genetic diversity could be explained by demographic variation had only little support, even in the native range. Except for a weak increase in neutral genetic diversity with density (see SI Appendix, Fig. S3f) and fecundity (see SI Appendix, Fig. S3g), there was little direct influence of demographic variables on genetic diversity. There are at least two explanations for this general lack of a demographic relationship. First, genetic structure can arise even under frequent dispersal[44]. Thus, although we found strong spatial genetic structure in the native range, it is possible that dispersal was high enough to mask any influence of demography on genetic diversity (the natural dispersal pathway, Fig. 1). Second, the fine scale of demographic sampling within sites (a few m^2) might not reflect effective population size[47]. This fits with our understanding of abiotic filters operating at all scales, while biotic filters, such as inter- and intra-specific interactions affecting demographic performance, generally operate at localised scales[10,13]. *Plantago lanceolata* is also highly genetically variable, within and outside its native range. Thus, the low power within sites might have limited our ability to draw conclusions about demographic influences on genetic diversity. Sampling more individuals per site in future might reveal stronger effects of fecundity, survival and population growth on genetic diversity.

In summary, genetic diversity in *P. lanceolata* appears to be shaped predominantly by temperature and precipitation gradients related to gene flow and admixture, rather than demographic variation. Our data support the prediction, that high dispersal would dilute demographic effects on genetic diversity (H1). Globally, our analyses suggest that genetic diversity in the non-native range is shaped by admixture from multiple source populations and ongoing introductions, leading to high neutral and adaptive genetic diversity (H2). Our data suggest that invasive populations can establish in a broad range of environments, without the need for associated demographic change. Thus, there was little support for the prediction that demographic variation could explain environmental effects on genetic diversity (H3). Our unique global demographic data set provides new evidence that invasive species can overcome ecological 'rules' in their non-native range[11-13]. Reducing long-distance dispersal and further introductions of invasive plants is important, even in areas where they already exist, as this will limit future increases in genetic diversity and the formation of new genotypes that confer an adaptive advantage in new environments.

Methods

Study overview

Plantago lanceolata is a short-lived (mean, max = 2.8, 8 yr[48]), perennial forb, native to Europe. It reproduces sexually and vegetatively, with gynodioecy, self-incompatibility and protogyny to enhance outcrossing[49]. Flowers are wind pollinated and seeds mature in summer. The species occurs in a wide range of habitats including semi-natural grasslands, roadsides, disturbed sites, abandoned fields and agricultural land[50]. Seeds are dispersed locally by wind but seed dispersal distances are estimated to be within centimetres or metres of the mother plant[51]. Widespread propagule movement by humans[29] and repeated introductions as seed contaminants[30] has led to the global distribution of *P. lanceolata*. It has been present in Australia since before 1850 (www.ala.org.au), in North America since before 1832[30] and for an unknown time in South Africa[52]. It is cultivated as a commercial pasture plant in New Zealand because it grows well in the mild winter and limits soil nitrification[31]. The species is classed as invasive in its non-native range[52] because it reproduces prolifically and spreads over large areas[53]. We follow this definition of 'invasive' to refer to *P. lanceolata* and other plant species with this characteristic. We use the term 'non-native' to refer to the geographic range outside of Europe where the species exists.

We used field-collected demographic and DNA data from populations of *P. lanceolata* to analyse spatial variation in demographic rates and genetic diversity. The demographic data were used to parameterise the simulation part of the study (H1) and to analyse the demographic influence on genetic diversity across global environmental gradients (H3). For the genetic data set, we sampled 454 individuals from 53 naturally-occurring populations in 21 countries across the native

European range (35 populations: Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Norway, Romania, Spain, Sweden, Switzerland, United Kingdom) and the non-native range (18 populations: Australia, Canada, Japan, New Zealand, South Africa, USA) (Fig. 2). The latitudinal range of sampling, in absolute terms, was 27.5–61.4°. Forty-four populations (83%) were established sites in the PLANTPOPNET network (www.plantpopnet.com) undergoing an annual demographic census, while the remaining nine were sampled for DNA only (see SI Appendix, Table S1).

We characterised the environment at each site using four variables from BioClim[54] at 30" resolution: annual mean temperature, annual mean precipitation, temperature seasonality (standard deviation of annual mean temperature) and precipitation seasonality (coefficient of variation in annual mean precipitation). We selected these variables because they were important for morphological variation in *P. lanceolata* in preliminary analyses and multi-collinearity was not high (variance inflation factor < 3, maximum *r* between pairs of environmental variables = 0.43 (mean temperature and seasonality in precipitation) and between range (native/non-native) and environment (mean temperature) = 0.59)[55].

Field demographic census & DNA sampling

PLANTPOPNET is an ongoing research project that began in 2014 and annual censuses of *P. lanceolata* populations are planned for the long-term. Our analysis used data collected between 2014 and 2017, but not all sites began data collection at the same time (i.e. year 0 varied among sites, see SI Appendix, Table S1). In most populations (61%), year 0 was 2015 and 73% of populations were sampled twice during this study period (number of annual censuses per population = 1–3, see SI Appendix, Table S1). At each census site in year 0, a series of adjacent 50 x 50 cm quadrats was established along transects until the quadrats covered 100 individual plants. Researchers established transects where *P. lanceolata* was present in sufficient numbers for demographic studies, so density estimates might reflect upper estimates across local populations. Quadrats were permanently marked to enable repeat censuses from year 1 onwards. Each plant was individually tagged and all rosettes on each plant were measured according to a standard protocol[56] which included leaf length, number of flowering stems, inflorescence length and stage of seed development.

At each site, fresh leaf tissue from seven to nine individuals was collected and placed immediately in silica gel (see SI Appendix, Table S1). Sampled individuals were close to (approximately 5–20 m), but outside of, census plots and were separated from each other by approximately 5–10 m. Thus, we avoided damage to permanently marked individuals in the census population, ensured that samples were closely related to the census population and minimised the chance of sampling clones. We included two samples each from one population of *P. coronopus* (Spain) and four populations of *P. major* (Australia x 2, Ireland x 1, Romania x 1) as outgroups. To investigate if naturally occurring populations were influenced by genetic stock from commercial pasture lines, we included nine individuals from each of three cultivar lines derived from *P. lanceolata*: AgriTonic, Ceres Tonic and Tonic Plantain. The whole data set thus included 491 individuals. The data are publicly available in Dataset S2 (https://doi.org/10.5281/zenodo.3579579).

Genotyping

Samples were genotyped at Diversity Arrays Technology P/L (Canberra, Australia) using double restriction enzyme complexity reduction and high-throughput sequencing (DArTseq). Total genomic DNA was extracted with a NucleoSpin 96 Plant II Core Kit (MACHEREY-NAGEL) and purified using a Zymo kit (Zymo Research). The enzymes PstI and MseI were chosen following tests of different enzyme combinations for *P. lanceolata*. DNA samples were processed in digestion / ligation reactions following Kilian et al.[57] but substituting the single PstI adaptor for two

adaptors corresponding to restriction enzyme-specific overhangs. The PstI adaptor was modified to include Illumina sequencing primers and variable length barcodes following Elshire et al.[58]. Mixed fragments (PstI-MseI) were amplified in 30 rounds of PCR using the following reaction conditions: 94 °C for 1 min, then 30 cycles of 94 °C for 20 sec, 58 °C for 30 sec, 72 °C for 45 sec, followed by 72 °C for 7 min. After PCR, equimolar amounts of amplification products from each sample were bulked and applied to c-Bot (Illumina) bridge PCR followed by single-read sequencing on an Illumina Hiseq2500 for 77 cycles. Raw sequences were processed using DArTseq analytical pipelines (DArTdb) to split samples by barcode and remove poor quality sequences. Genotypes for co-dominant, single nucleotide polymorphisms (SNPs) were called de novo (i.e. without a reference genome) from 69 bp sequences using DArTseq proprietary software (DArTsoft). Replicate samples were processed to assess call rate (mean = 79%), reproducibility (mean = 99 %) and polymorphic information content (mean = 22%).

SNP filtering

Starting with 37,692 SNPs that passed DArTseq quality control, we filtered the data for minimum minor allele frequency (1%), call rate (50%) and reproducibility (98%) using custom R scripts[59] (Dataset S2). Loci in Hardy-Weinberg (HW) and linkage disequilibrium hold important biological information about population structure but extreme disequilibrium can indicate genotyping errors which bias estimates of population structure[60]. Within sites, there was limited power to reliably test for patterns of HW and linkage disequilibrium (7–9 individuals per site). It was not possible to combine samples from multiple populations because we detected strong genetic structure, even within countries, which would have produced biologically meaningful patterns of disequilibrium arising from the Wahlund effect[61]. Thus, to identify SNPs with consistent patterns of HW disequilibrium, we tested each locus in every population separately using Fisher's exact tests[62] and used un-adjusted *P* values given the low power within sites. Loci which deviated from HW equilibrium in > 5 populations were removed[63]. We used the correlation between genotype frequencies[64] to test for linkage disequilibrium between each pair of loci in each population. Following the same rationale as for HW disequilibrium, we removed a locus if it was in a correlated pair ($r > 0.75$) in > 5 populations. To reduce the chance of disequilibrium from physical linkage, we also filtered SNPs that occurred in the same 69 bp sequence as another SNP, keeping the one with the highest call rate. The data comprised 21,190 SNPs after applying these filters.

Detecting loci under putative selection

Neutrality was an assumption underlying the population structure models we used, thus, we investigated if SNPs were putatively under selection using one population-level method (BayeScan) and two individual-level methods (PCAdapt and LFMM). BayeScan uses an MCMC algorithm to examine outlier loci against background values of population differentiation (F_{ST}) among pre-defined populations[65]. PCAdapt and LFMM both define background population structure as *K* principal components derived from individual genotypes[66,67]. In PCAdapt, each SNP is regressed against each principal component. LFMM uses the principal components as latent factors in a Gaussian mixed model, where the genotype matrix is modelled as a function of an environmental matrix[67]. While BayeScan is suitable for our population-level sampling design, PCAdapt and LFMM are more reliable for species with complex, hierarchical population structure (e.g. multiple divergence events) and are less sensitive to admixed individuals and outliers in the data[68,69]. Thus, we considered outliers identified in any of the three methods to be putatively under selection.

For BayeScan, we set the prior odds at 200 (appropriate for the number of markers in our data[70]), ran the model using default parameters (100,000 iterations with a thinning interval of 10, a burn-in of 50,000 and 20 pilot runs of 5,000 iterations), and checked the distribution of the log likelihood across iterations to ensure model convergence (see SI Appendix, Fig. S5). For both individual-

level methods, we examined scree plots to determine *K* and used the first 10 components which captured the majority of population structure in the data (see SI Appendix, Fig. S5). We defined the LFMM environmental matrix using the four 30" BioClim variables described above and three additional variables: elevation (metres above sea level, measured at the site) and two variables extracted from CliMond[71] at 5' resolution: annual mean moisture index and seasonality in moisture (CV of annual mean moisture). To control for false discovery rate, we calculated q-values from p-values and classed SNPs as outliers where q < 0.05 for BayeScan and PCAdapt and q < 0.1 for LFMM (to account for the small number of loci identified with this method, see SI Appendix, Fig. S5). The three analyses identified a total of 3,026 outlier SNPs and, as commonly reported in other studies[69], there was little overlap among methods (see SI Appendix, Fig. S5). After filtering the putatively adaptive loci, our final data set comprised 18,164 neutral SNPs.

Simulated genetic diversity (Hypothesis 1)

We conducted two simulation experiments in MetaPopGen 0.0.4[72] to determine if realistic levels of variation in *P. lanceolata* survival and fecundity would influence genetic diversity and whether dispersal would override demographic influences on genetic diversity. Gametes in the model are produced via Mendelian segregation and mating is random[72]. We modelled two distinct populations to examine different rates of juvenile survival and female fecundity. In Experiment 1, the two populations were unconnected by dispersal, while in Experiment 2 they were connected by varying levels of dispersal.

Male fecundity δ_M in *P. lanceolata* is high (10,000–54,000 pollen grains per anther[32]) and had no influence on genetic diversity. Thus, we set δ_M at 10,000 and focussed on variation in female fecundity (seeds per plant) δ_F, adult σ_a and juvenile σ_j survival rate, and between-population dispersal δ (number of migrants per generation). In both experiments each of the two populations *i*, had two age classes *x* (juvenile x_j, adult x_a), three genotypes *p* representing all combinations of two alleles (00, 01 and 11) and a starting size N_{xp} of 25,000 individuals. The model was not spatially explicit, but we wanted each population to represent a 1 ha site with a density of 15 individuals / m^2 (based on census data from year 0). Generation time in *P. lanceolata* is approximately 3 years (range 1–3 years[73,74]). Thus, we ran the model for 100 time steps to represent population dynamics over 100–300 years, accounting approximately for the time *P. lanceolata* has been present in its non-native range. Population sizes reached a steady state within 10 time steps. We estimated juvenile carrying capacity as $K = (\delta_F * (N*p)) * g$, where *g* is the estimated field germination rate (0.039). We kept *K* time- and population-constant. MetaPopGen can only simulate one locus at a time, so we repeated the experiments 300 times to simulate sampling 300 independent loci (following[72]).

In Experiment 1, we tested the influence of δ_F on genetic diversity (1–100, based on census data from year 0) and σ ($\sigma_{ji1} = 0.1$; $\sigma_{ai1} = 0.84$; $\sigma_{ji2} = 0.2$; $\sigma_{ai2} = 0.71$) with no dispersal between populations (δ=0). Survival rates were based on a total population estimate of 5% alive after five years (exp(log(0.05)/5)) (ref. [73]) and adjusted for commonly reported low survival in juveniles[42]. In Experiment 2, we tested the influence of δ (migration rate: 0–0.04 = number of migrants: 0–60,000) on the difference in genetic diversity between populations. Each population had the same survival rates as Experiment 1 and δ_F was kept constant at 20. The migration rates produce large numbers of migrants because each plant produces 20 'newborns' and migration occurs before recruitment in the model[72]. Thus, δ is influenced by *K* and will always be higher than recruitment. We summarised expected heterozygosity at the end of each simulation and calculated the mean and 95% confidence interval across the 300 loci. The experiments can be reproduced with the code in Dataset S2.

Population genetic structure (Hypothesis 2)

All population structure analyses used our panel of neutral SNPs; a choice dictated by the model assumptions being based on Hardy-Weinberg and linkage equilibrium. We first conducted an Analysis of Molecular Variance in poppr 2.8.0[75] to determine how neutral genetic diversity was partitioned across levels: within individuals, among individuals within populations, among populations within ranges, and between the native and non-native range. To assess genomic relationships and the degree of admixture in the global data set, we used fastSTRUCTURE[34]. This model determines the number of genetic clusters in the data that would maximise Hardy-Weinberg and linkage equilibrium (K). We investigated K=1 to K=20 and assigned each individual to a cluster based on the model complexity that maximised marginal likelihood and the model components used to explain structure in data[34]. To quantify the level of admixture for each individual (i) across the most likely K, we calculated a Diversity Score[76] as:

$$DS = \frac{-\sum_{i=1}^{K} C_i \cdot \ln(C_i)}{-H_{\max}}$$

where C_i is the cumulative admixture and $H_{\max}$ is a scaling factor ($H_{\max} = K \cdot ((1/K) \cdot \ln(1/K))$), making DS relative to complete evenness for each individual. We used a linear mixed model to evaluate whether there was a difference in DS between the native and non-native range, with site fitted as a random effect.

To determine whether multiple introductions of *P. lanceolata* had occurred in non-native regions (Australia, Japan, New Zealand, North America and South Africa) we estimated the minimum number of propagules required to produce the observed level of genetic diversity in non-native regions ($Prop_{min}$)[77]. We defined the source population as all of Europe because non-native individuals were usually composed of admixed genotypes from multiple European populations. For each non-native region, we calculated the number of alleles not present in Europe and removed these from the reference panel of non-native alleles. Individuals from the native range were then randomly cumulatively sampled without replacement. $Prop_{min}$ was the number of individuals sampled at the point when all alleles in the non-native panel were represented (Dataset S2). We repeated the process 1000 times to obtain a mean and standard error. We also calculated the number of unique alleles in each of the 53 sites as a measure of uniqueness.

To assess the influence of environmental gradients on spatial genetic structure, we used generalised dissimilarity models[78,79]. We fitted one model for the native range, a second for the non-native range and a third for the global data set (native and non-native). We calculated genetic differentiation as F_{ST} between all pairs of populations in GENEPOP 4.6[80]. Environmental distances between all pairs of populations i and j were calculated from the four BioClim variables x ($x_i - x_j$)[79]. For each of the three data sets, we fitted geographic distance and all environmental distances as predictor variables in a single model. The importance of each variable, given all other variables, was assessed by comparing the fitted model to 500 models with a permuted environmental matrix[79]. Thus, the effect of each environmental variable can be interpreted independently and differences in spatial scale are accounted for by the geographic distance variable. P values were Bonferroni-adjusted across all terms within each model. We used deviance explained to assess goodness-of-fit of the three models. Given samples size differences between the three data sets, we used a bootstrap estimate from 10,000 replicates of the deviance explained to assess the accuracy of the model fit. We assumed the deviance explained to be accurate if bootstrap 95% confidence interval (CI) did not include zero.

Demographic & dispersal effects on genetic diversity (Hypothesis 3)

We used linear regression to determine if environmental influences on within-population genetic diversity could be explained by demographic variation and whether this effect would be weakened

by mass dispersal into the non-native range (Hypothesis 3). The observation-level for all analyses was the population and the number of observations was 44 (i.e. all populations with genetic and demographic data, see SI Appendix, Table S1).

Genetic diversity was calculated as allelic richness in hierfstat[81], separately for the neutral (18,166 SNPs) and adaptive (3,024 SNPs) datasets. Allelic richness was highly correlated with expected heterozygosity (H_e) (r = 0.98) and, because it was standardised for sample size, it eliminated a weak correlation we observed between H_e and sample size. We characterised the environment using the four BioClim variables. For demography, we used three variables that can influence genetic diversity (Table 1): population density (rosettes/m^2), fecundity and empirical population growth rate. For fecundity, we used reproductive effort, estimated as the rosette-level inflorescence length x number of flowering stems per m^2. Empirical population growth rate was calculated as $r = \log(N_{t+1}/N_t)$, indicating the strength and direction of change in rosettes/m^2 in the first two years of the study (for 38 of the 44 populations with two years of data, see SI Appendix, Table S1). Thus, r reflects the combined influence of fecundity and survival (the variables explored in simulation Experiment 1). We used rosette-level data for all metrics to reduce potential observer bias in assessing clonality, but plant- and rosette-level metrics were highly correlated (r = 0.94). Fecundity was log-transformed to address a strongly skewed distribution and all predictors were standardised prior to analysis (x − mean(x)/SD(x)).

We tested environmental and demographic effects separately, to determine which variables best described variation in genetic diversity. The analysis comprised two stages. First, we analysed the effect of each environmental variable on genetic diversity. Here, we also modelled the environmental effect on demography (i.e. using the three demographic variables as response terms) to establish a baseline for environmental influences on demographic rates. Second, we examined whether each demographic variable influenced genetic diversity. In both stages we analysed environmental and demographic interactions with range (native/non-native). Because data limitations (n = 44) it was not possible to fit complex models with multiple interaction terms so we modelled each predictor separately.

To determine the importance of each environmental or demographic predictor, we used AICc to compare model fit across five alternative model forms: a null model (no predictor variation), a predictor only model, a range only model, an additive model (predictor + range) and an interactive model (predictor x range). We considered a model to have support from the data if it improved the fit over the null model by ΔAICc > 2 (ref. [82]). Among models that out-fitted the null, those within ΔAICc ≤ 2 of each other were considered to have equal support from the data. In these cases, we presented the top-ranked model in the main document and supported models in the Supporting Information. To interpret interaction models in light of sample size differences between the native (30) and non-native (14) ranges (e.g. a strong response in the native range and no response in the non-native range), we obtained a bootstrap 95% confidence interval (CI) from 10,000 bootstrap replicates of the interaction coefficient using the adjusted bootstrap percentile method.

Acknowledgments

Jan van Groenendael helped design the PLANTPOPNET network. Leander Anderegg, Lauchlan Fraser, Jennifer Gremer, Emily Griffoul, Adrian Oprea, Richard Shefferson and Danielle Sherman provided data. Maeve Harrison assisted with field work. Valuable discussions with Alan Stewart and Andrzej Kilian improved our knowledge of *Plantago* cultivation and SNP data generation, respectively.

References

1. Smith, A. L. *et al.* Dispersal responses override density effects on genetic diversity during post-disturbance succession. *Proceedings of the Royal Society of London B* **283**, 20152934 (2016).

2. Duminil, J. *et al.* Can population genetic structure be predicted from life-history traits? *Am. Nat.* **169**, 662-672 (2007).
3. Leffler, E. M. *et al.* Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol.* **10**, e1001388 (2012).
4. Ellegren, H. & Galtier, N. Determinants of genetic diversity. *Nat. Rev. Genet.* **17**, 422–433 (2016).
5. Romiguier, J. *et al.* Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* **515**, 261-263 (2014).
6. Leimu, R., Mutikainen, P. I. A., Koricheva, J. & Fischer, M. How general are positive relationships between plant population size, fitness and genetic variation? *J. Ecol.* **94**, 942-952 (2006).
7. Gaggiotti, O. E. Population genetic models of source–sink metapopulations. *Theor. Popul. Biol.* **50**, 178-208 (1996).
8. Hughes, A. R., Inouye, B. D., Johnson, M. T. J., Underwood, N. & Vellend, M. Ecological consequences of genetic diversity. *Ecol. Lett.* **11**, 609-623 (2008).
9. Vergeer, P., Rengelink, R., Copal, A. & Ouborg, N. J. The interacting effects of genetic variation, habitat quality and population size on performance of *Succisa pratensis*. *J. Ecol.* **91**, 18-26 (2003).
10. van Kleunen, M., Bossdorf, O. & Dawson, W. The ecology and evolution of alien plants. *Annual Review of Ecology, Evolution, and Systematics* **49**, 25-47 (2018).
11. Catford, J. A., Bode, M. & Tilman, D. Introduced species that overcome life history tradeoffs can cause native extinctions. *Nature Communications* **9**, 2131 (2018).
12. van Boheemen, L. A., Atwater, D. Z. & Hodgins, K. A. Rapid and repeated local adaptation to climate in an invasive plant. *New Phytol.* **222**, 614-627 (2019).
13. Endriss, S. B., Alba, C., Norton, A. P., Pyšek, P. & Hufbauer, R. A. Breakdown of a geographic cline explains high performance of introduced populations of a weedy invader. *J. Ecol.* **106**, 699-713 (2017).
14. Dlugosch, K. M. & Parker, I. M. Invading populations of an ornamental shrub show rapid life history evolution despite genetic bottlenecks. *Ecol. Lett.* **11**, 701-709 (2008).
15. Wilson, J. R. U., Dormontt, E. E., Prentis, P. J., Lowe, A. J. & Richardson, D. M. Something in the way you move: dispersal pathways affect invasion success. *Trends Ecol. Evol.* **24**, 136-144 (2009).
16. Estoup, A. *et al.* Is there a genetic paradox of biological invasion? *Annual Review of Ecology, Evolution, and Systematics* **47**, 51-72 (2016).
17. Rius, M. & Darling, J. A. How important is intraspecific genetic admixture to the success of colonising populations? *Trends Ecol. Evol.* **29**, 233-242 (2014).
18. Parepa, M., Fischer, M., Krebs, C. & Bossdorf, O. Hybridization increases invasive knotweed success. *Evolutionary Applications* **7**, 413-420 (2014).
19. Exposito-Alonso, M. *et al.* The rate and potential relevance of new mutations in a colonizing plant lineage. *PLoS Genetics* **14**, e1007155 (2018).
20. Dlugosch, K. M., Anderson, S. R., Braasch, J., Cang, F. A. & Gillette, H. D. The devil is in the details: genetic variation in introduced populations and its contributions to invasion. *Mol. Ecol.* **24**, 2095-2111 (2015).
21. Crawford, K. M. & Whitney, K. D. Population genetic diversity influences colonization success. *Mol. Ecol.* **19**, 1253-1263 (2010).
22. Oduor, A. M. O., Leimu, R. & van Kleunen, M. Invasive plant species are locally adapted just as frequently and at least as strongly as native plant species. *J. Ecol.* **104**, 957-968 (2016).
23. Parker, J. D. *et al.* Do invasive species perform better in their new ranges? *Ecology* **94**, 985-994 (2013).
24. Uesugi, A. & Kessler, A. Herbivore exclusion drives the evolution of plant competitiveness via increased allelopathy. *New Phytol.* **198**, 916-924 (2013).
25. Arredondo, T. M., Marchini, G. L. & Cruzan, M. B. Evidence for human-mediated range expansion and gene flow in an invasive grass. *Proceedings of the Royal Society B: Biological Sciences* **285**, 20181125 (2018).
26. Salguero-Gómez, R. *et al.* The COMPADRE Plant Matrix Database: an open online repository for plant demography. *J. Ecol.* **103**, 202-218 (2015).
27. Li, S.-L., Vasemägi, A. & Ramula, S. Genetic variation facilitates seedling establishment but not population growth rate of a perennial invader. *Annals of Botany* **117**, 187-194 (2016).
28. Keller, S. R., Fields, P. D., Berardi, A. E. & Taylor, D. R. Recent admixture generates heterozygosity–fitness correlations during the range expansion of an invading species. *J. Evol. Biol.* **27**, 616-627 (2014).
29. Pickering, C. & Mount, A. Do tourists disperse weed seed? A global review of unintentional human-mediated terrestrial seed dispersal on clothing, vehicles and horses. *Journal of Sustainable Tourism* **18**, 239-256 (2010).

30. Mack, R. N. & Erneberg, M. The United States naturalized flora: largely the product of deliberate introductions. *Annals of the Missouri Botanical Garden* **89**, 176-189 (2002).
31. Skinner, R. H. & Stewart, A. V. Narrow-leaf plantain (*Plantago lanceolata* L.) selection for increased freezing tolerance. *Crop Science* **54**, 1238-1242 (2014).
32. Primack, R. B. Evolutionary aspects of wind pollination in the genus *Plantago* (Plantaginaceae). *New Phytol.* **81**, 449-458 (1978).
33. Pahl, A. T., Kollmann, J., Mayer, A. & Haider, S. No evidence for local adaptation in an invasive alien plant: field and greenhouse experiments tracing a colonization sequence. *Annals of Botany* **112**, 1921-1930 (2013).
34. Raj, A., Stephens, M. & Pritchard, J. K. fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics* **197**, 573-589 (2014).
35. Rieseberg, L. H., Widmer, A., Arntz, A. M. & Burke, B. The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. *Philosophical Transactions of the Royal Society B: Biological Sciences* **358**, 1141-1147 (2003).
36. Fenollosa, E., Roach, D. A. & Munné-Bosch, S. Death and plasticity in clones influence invasion success. *Trends Plant Sci.* **21**, 551-553 (2016).
37. François, O., Blum, M. G. B., Jakobsson, M. & Rosenberg, N. A. Demographic history of European populations of *Arabidopsis thaliana*. *PLOS Genetics* **4**, e1000075 (2008).
38. Pinhasi, R., Fort, J. & Ammerman, A. J. Tracing the origin and spread of agriculture in Europe. *PLoS Biol.* **3**, e410 (2005).
39. Grosvenor, M. J. *et al.* Human activity was a major driver of the mid-Holocene vegetation change in southern Cumbria: implications for the elm decline in the British Isles. *Journal of Quaternary Science* **32**, 934-945 (2017).
40. Kajtoch, Ł. *et al.* Phylogeographic patterns of steppe species in Eastern Central Europe: a review and the implications for conservation. *Biodivers. Conserv.* **25**, 2309-2339 (2016).
41. Villellas, J., Doak, D. F., García, M. B. & Morris, W. F. Demographic compensation among populations: what is it, how does it arise and what are its implications? *Ecol. Lett.* **18**, 1139-1152 (2015).
42. Roach, D. A. Age-specific demography in *Plantago*: variation among cohorts in a natural plant population. *Ecology* **84**, 749-756 (2003).
43. Villellas, J., Berjano, R., Terrab, A. & García, M. B. Escasa correspondencia entre diversidad genética y demografía en una planta a escala continental. *Ecosistemas* **28**, 4-14 (2019).
44. Fraser, C. I., Davies, I. D., Bryant, D. & Waters, J. M. How disturbance and dispersal influence intraspecific structure. *J. Ecol.* **106**, 1298-1306 (2018).
45. Ohbayashi, K., Hodoki, Y., I. Kondo, N., Kunii, H. & Shimada, M. A massive tsunami promoted gene flow and increased genetic diversity in a near threatened plant species. *Scientific Reports* **7**, 10933 (2017).
46. Vallejo-Marín, M., Dorken, M. E. & Barrett, S. C. H. The ecological and evolutionary consequences of clonality for plant mating. *Annual Review of Ecology, Evolution, and Systematics* **41**, 193-213 (2010).
47. Coutts, S. R., Salguero-Gómez, R., Csergő, A. M. & Buckley, Y. M. Extrapolating demography with climate, proximity and phylogeny: approach with caution. *Ecol. Lett.* **19**, 1429-1438 (2016).
48. Roeder, A., Schweingruber, F. H., Fischer, M. & Roscher, C. Growth ring analysis of multiple dicotyledonous herb species—A novel community-wide approach. *Basic Appl. Ecol.* **21**, 23-33 (2017).
49. Krohne, D. T., Baker, I. & Baker, H. G. The maintenance of the gynodioecious breeding system in *Plantago lanceolata* L. *The American Midland Naturalist* **103**, 269-279 (1980).
50. Sagar, G. & Harper, J. Biological flora of the British Isles. *Plantago major* L., *P. media* L. and *P. lanceolata*. *J. Ecol.* **52**, 189-221 (1964).
51. Tonsor, S. J. Leptokurtic pollen-flow, non-leptokurtic gene-flow in a wind-pollinated herb, *Plantago lanceolata* L. *Oecologia* **67**, 442-446 (1985).
52. Alston, K. P. & Richardson, D. M. The roles of habitat features, disturbance, and distance from putative source populations in structuring alien plant invasions at the urban/wildland interface on the Cape Peninsula, South Africa. *Biol. Conserv.* **132**, 183-198 (2006).
53. Richardson, D. M. *et al.* Naturalization and invasion of alien plants: concepts and definitions. *Divers. Distrib.* **6**, 93-107 (2000).
54. Booth, T. H., Nix, H. A., Busby, J. R. & Hutchinson, M. F. BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. *Divers. Distrib.* **20**, 1-9 (2013).
55. Zuur, A. F., Ieno, E. N. & Elphick, C. S. A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evol.* **1**, 3-14 (2010).

56. Buckley, Y. M. *et al.* Plantpopnet protocol V1.01 2015. Figshare Documen, doi.org/10.6084/m6089.figshare.7982810 (2019).
57. Kilian, A. *et al.* Diversity arrays technology: a generic genome profiling technology on open platforms. *Data Production and Analysis in Population Genomics: Methods and Protocols*, 67-89 (2012).
58. Elshire, R. J. *et al.* A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**, e19379 (2011).
59. R: a language and environment for statistical computing. (http://www.R-project.org, R Foundation for Statistical Computing, Vienna, Austria., 2017).
60. Turner, S. *et al.* Quality control procedures for genome wide association studies. *Current Protocols in Human Genetics* **68**, 1.19.11–11.19.18 (2012).
61. Slatkin, M. Linkage disequilibrium – understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet.* **9**, 477-485 (2008).
62. Wigginton, J. E., Cutler, D. J. & Abecasis, G. R. A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics* **76**, 887-893 (2005).
63. Schilling, M. P. *et al.* Genotyping-by-sequencing for *Populus* population genomics: an assessment of genome sampling patterns and filtering approaches. *PLOS ONE* **9**, e95292 (2014).
64. Chan, E. *Handy R functions for genetics research*, https://github.com/ekfchan/evachan.org-Rscripts (2018).
65. Foll, M. & Gaggiotti, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977-993 (2008).
66. Duforet-Frebourg, N., Bazin, E. & Blum, M. G. Genome scans for detecting footprints of local adaptation using a Bayesian factor model. *Molecular Biology and Evolution* **31**, 2483-2495 (2014).
67. Frichot, E., Schoville, S. D., Bouchard, G. & François, O. Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution* **30**, 1687-1699 (2013).
68. Luu, K., Bazin, E. & Blum, M. G. pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Mol. Ecol. Res.* **17**, 67-77 (2017).
69. Villemereuil, P., Frichot, É., Bazin, É., François, O. & Gaggiotti, O. E. Genome scan methods against more complex models: when and how much should we trust them? *Mol. Ecol.* **23**, 2006-2019 (2014).
70. Foll, M. *BayeScan v2.1 user manual*, http://cmpg.unibe.ch/software/BayeScan/ (2012).
71. Kriticos, D. J. *et al.* CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods Ecol. Evol.* **3**, 53-64 (2011).
72. Andrello, M. & Manel, S. MetaPopGen: an r package to simulate population genetics in large size metapopulations. *Mol. Ecol. Res.* **15**, 1153-1162 (2015).
73. van Groenendael, J. M. & Slim, P. The contrasting dynamics of two populations of *Plantago lanceolata* classified by age and size. *J. Ecol.* **76**, 585-599 (1988).
74. Steiner, U., Tuljapurkar, S. & Roach, D. Quantifying genetic, environmental and individual stochastic variability in *Plantago lanceolata*. *bioRxiv*, 270603 (2018).
75. Kamvar, Z. N., Brooks, J. C. & Grünwald, N. J. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics* **6**, 208-208 (2015).
76. Harismendy, O., Kim, J., Xu, X. & Ohno-Machado, L. Evaluating and sharing global genetic ancestry in biomedical datasets. *Journal of the American Medical Informatics Association: JAMIA* **26**, 457-461 (2019).
77. Alsos, I. G. *et al.* Frequent long-distance plant colonization in the changing arctic. *Science* **316**, 1606-1609 (2007).
78. Fitzpatrick, M. C. & Keller, S. R. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecol. Lett.* **18**, 1-16 (2015).
79. Ferrier, S., Manion, G., Elith, J. & Richardson, K. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Divers. Distrib.* **13**, 252-264 (2007).
80. Rousset, F. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Res.* **8**, 103-106 (2008).
81. Goudet, J. & Jombart, T. *hierfstat: estimation and tests of hierarchical F-statistics. R package version 0.04-22.*, 2015).
82. Hegyi, G. & Garamszegi, L. Using information theory as a substitute for stepwise regression in ecology and behavior. *Behav. Ecol. Sociobiol.* **65**, 69-76 (2011).

83. Pannell, J. R. & Charlesworth, B. Effects of metapopulation processes on measures of genetic diversity. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **355**, 1851-1864 (2000).
84. *GBIF Secretariat* (Checklist dataset https://doi.org/10.15468/39omei, accessed via GBIF.org, 2016).
85. Enquist, B. J., Condit, R., Peet, R. K., Schildhauer, M. & Thiers, B. M. Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity. *PeerJ Preprints* **4**, e2615v2612 (2016).

Figures and Tables

Figure 1. Conceptual diagram showing how demographic performance and dispersal collectively shape genetic diversity in plant populations (+ indicates a positive relationship expected). Genetic diversity is influenced through natural pathways (solid line), such as local environmental conditions which affect demographic performance and effective population size[4]. Environmental conditions also affect genetic diversity though dispersal (e.g. by facilitating dispersal vectors or creating dispersal barriers). Dispersal can increase genetic diversity directly by providing a source of new genetic material (outcrossing) or indirectly through immigration and consequent effects on demography. High propagule pressure arising from high fecundity can influence source-sink dynamics[7,83], increasing rates of dispersal (hence the double arrow between demography and dispersal). Human activity can affect genetic diversity (dashed lines) by altering environmental conditions (e.g. climate change) and by changing dispersal rates and dispersal pathways (e.g. admixture). When this occurs, demographic performance can also be affected (e.g. through enemy release associated with dispersal across biogeographic boundaries) which can cause invasive plants to overcome biotic constraints on life-history[11] and environment-trait relationships[13]. Although genetic architecture can influence demography and dispersal, the overall quantity of neutral genetic diversity across the genome is more likely to be the outcome of demographic and dispersal processes, hence the one-sided arrows between these panels.

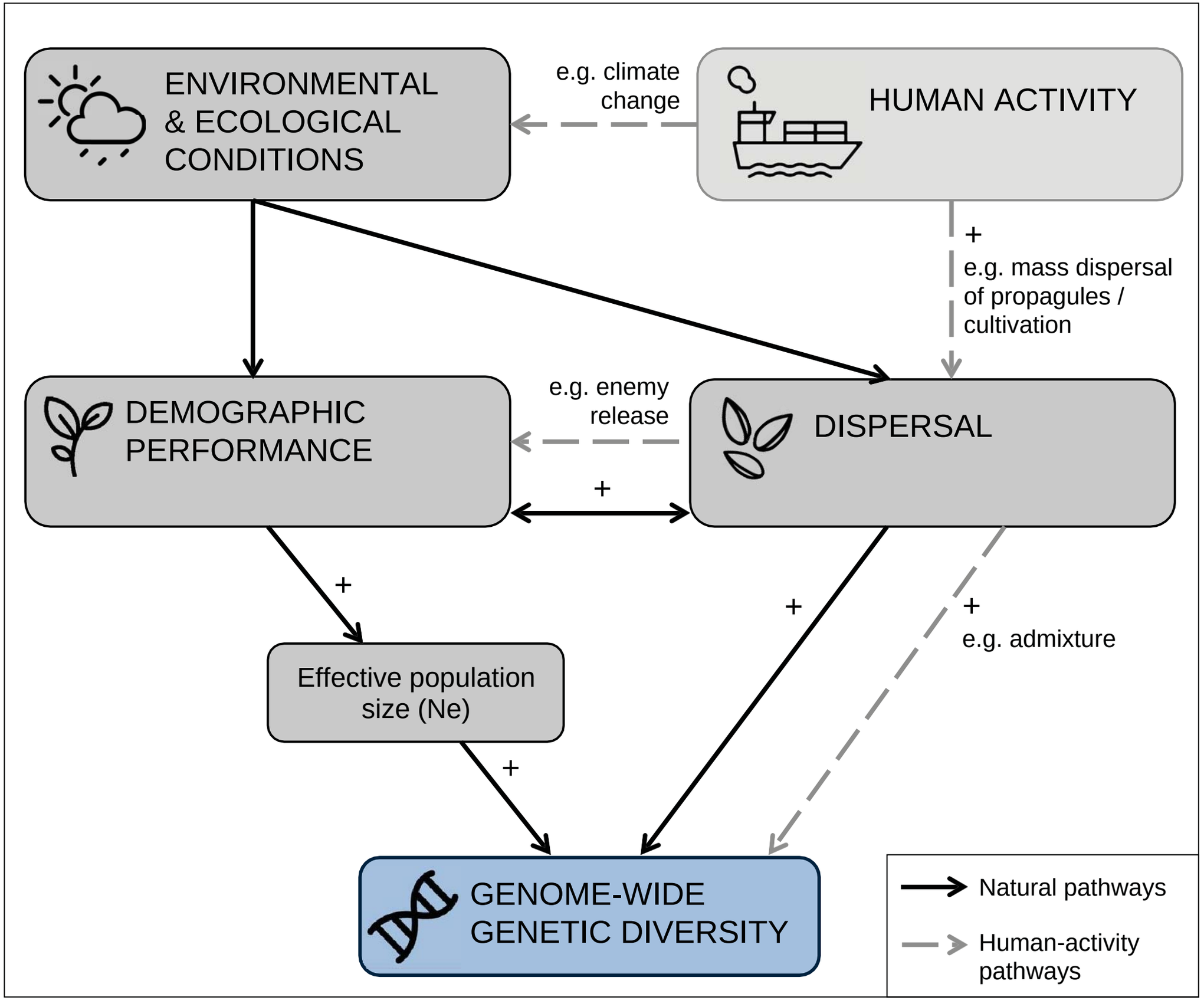

ENVIRONMENTAL & ECOLOGICAL CONDITIONS
e.g. climate change
HUMAN ACTIVITY
+
e.g. mass dispersal of propagules / cultivation
DEMOGRAPHIC PERFORMANCE
e.g. enemy release
DISPERSAL
+
+
Effective population size (Ne)
+
+
+
e.g. admixture
GENOME-WIDE GENETIC DIVERSITY
Natural pathways
Human-activity pathways

Figure 2. Global genetic structure in *Plantago lanceolata*. (a) Coloured bars represent the proportion of individual genotypes in each population assigned to one of six genetic clusters identified with fastSTRUCTURE. For clarity, multiple sites were aggregated where overlapping bars had similar assignment probabilities (e.g. southern Ireland, Switzerland). Dark grey points are *P. lanceolata* records from GBIF/BIENGBIF [84,85]. For each non-native region, the minimum number of propagules (mean ± standard error), overall ($Prop_{min}$) and relative to sample size ($Prop_{min}$ / N), indicates that multiple introductions would be required to produce observed levels of genetic diversity. The number of non-European alleles indicates that more genetic diversity was present in non-native regions than could be explained by the native sample. (b) Probability of assignment for 491 individuals to six genetic clusters, with individuals grouped by population within region. Three commercial cultivar lines and two outgroups (*P. coronopus* and *P. major*) were included. Country codes for each population are shown on the x-axis.

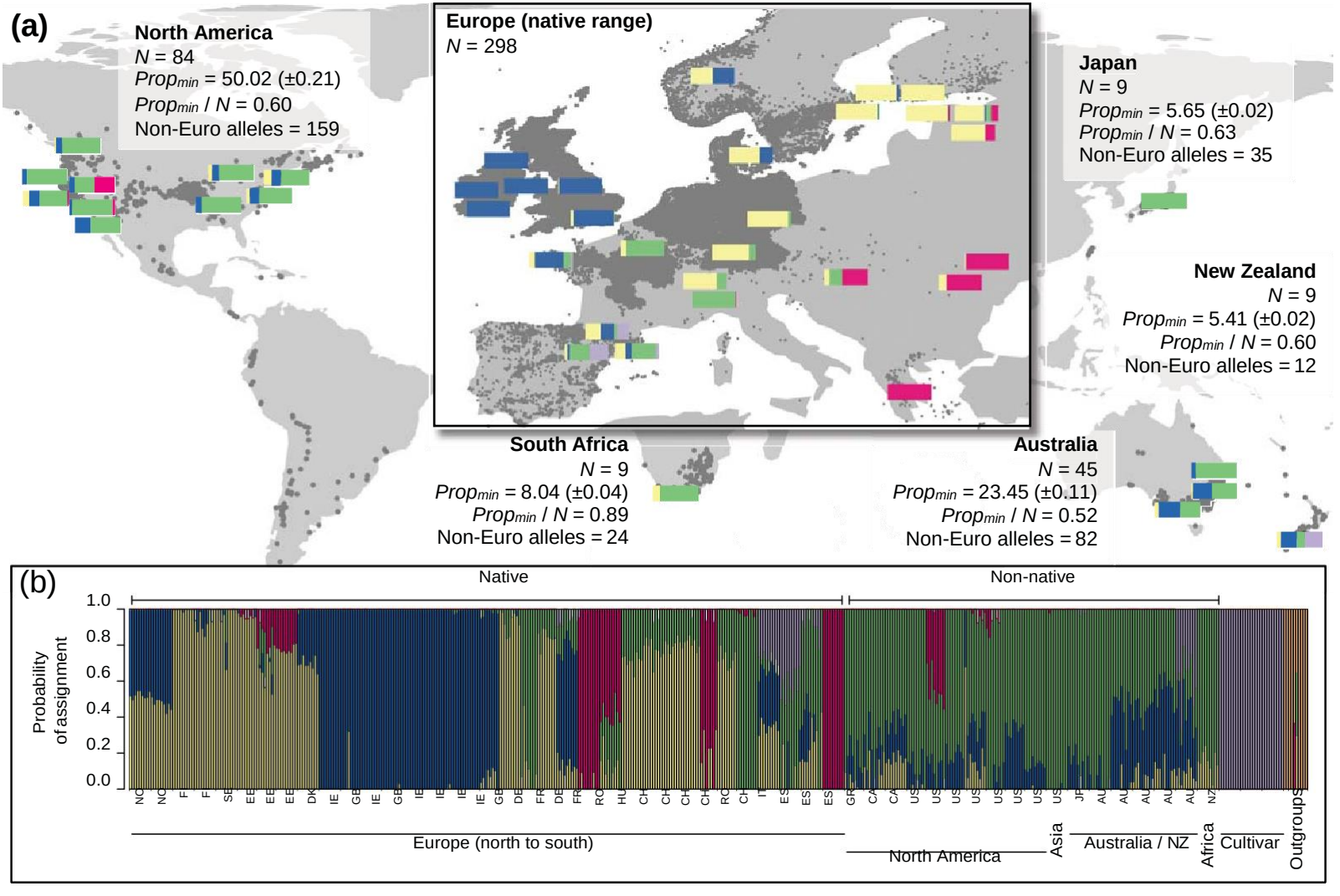
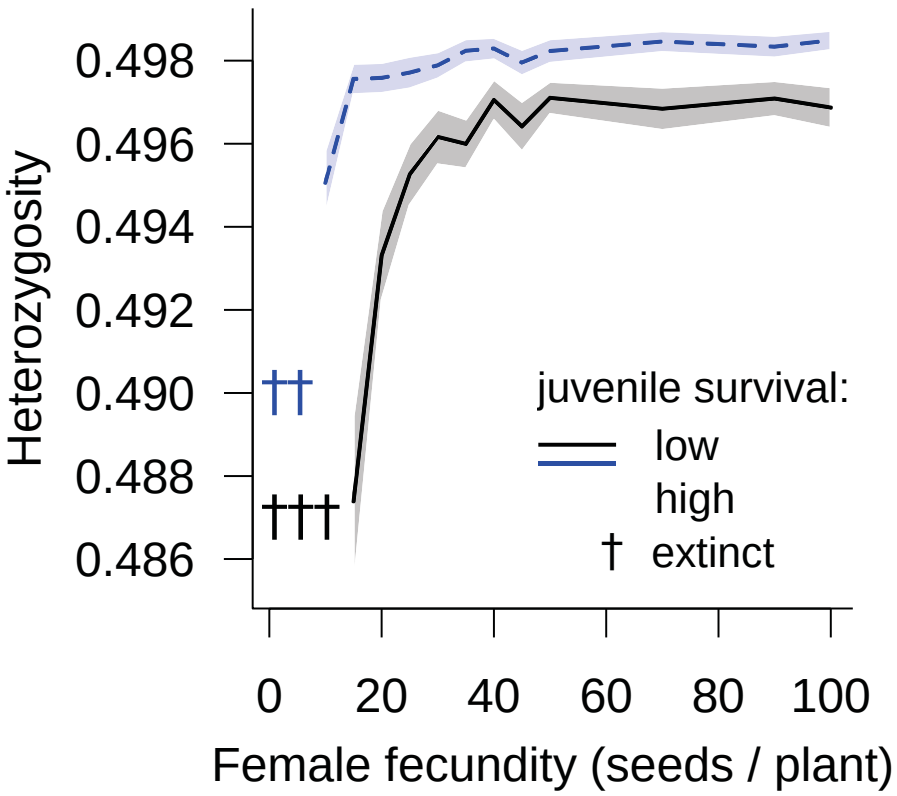
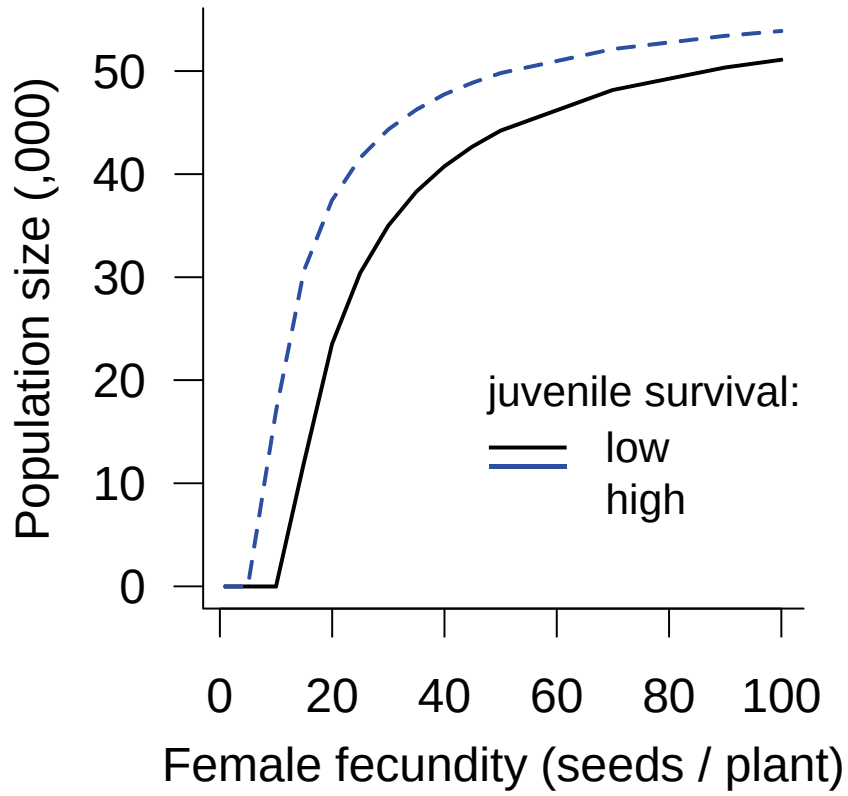

(a)
North America
N = 84
Prop_min = 50.02 (±0.21)
Prop_min / N = 0.60
Non-Euro alleles = 159
Europe (native range)
N = 298
Japan
N = 9
Prop_min = 5.65 (±0.02)
Prop_min / N = 0.63
Non-Euro alleles = 35
New Zealand
N = 9
Prop_min = 5.41 (±0.02)
Prop_min / N = 0.60
Non-Euro alleles = 12
South Africa
N = 9
Prop_min = 8.04 (±0.04)
Prop_min / N = 0.89
Non-Euro alleles = 24
Australia
N = 45
Prop_min = 23.45 (±0.11)
Prop_min / N = 0.52
Non-Euro alleles = 82
(b)
Native
Non-native
Probability of assignment
1.0
0.8
0.6
0.4
0.2
0.0
Europe (north to south)
North America
Asia
Australia / NZ
Africa
Cultivar
Outgroups

Figure 3. The simulated effect of demography and dispersal on genetic diversity (expected heterozygosity, ± 95% confidence interval) in two populations of *Plantago lanceolata*. (a) When there was no dispersal between populations, the population with high juvenile survival (o_j = 0.2) had greater genetic diversity than the population with low juvenile survival (o_j = 0.1). At very low levels of female fecundity δ_F, populations went extinct (†) but δ_F had little influence on genetic diversity at approximately > 25 seeds per plant. (b) Variation in o_j influenced population size at the end of the simulation. (c) The difference in heterozygosity between the two populations was influenced by dispersal between them (where fecundity was kept constant at 20 seeds / plant). (d) Genetic differences persisted until high levels of dispersal (> 50,000 migrants per generation) indicated by the 95% confidence interval crossing zero.

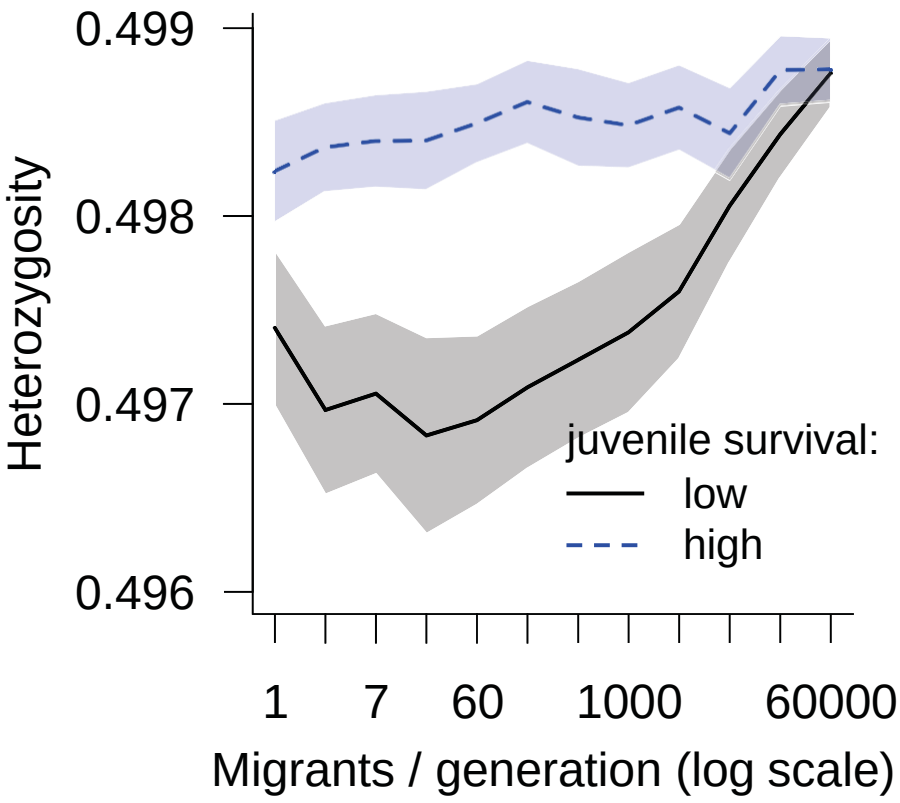
(a)


no dispersal
Heterozygosity
0.498
0.496
0.494
0.492
0.490
0.488
0.486
††
†††
juvenile survival:
low
high
† extinct
0
20
40
60
80
100
Female fecundity (seeds / plant)

(b)


no dispersal
Population size (,000)
50
40
30
20
10
0
juvenile survival:
low
high
0
20
40
60
80
100
Female fecundity (seeds / plant)

(c)


varying dispersal
Heterozygosity
0.499
0.498
0.497
0.496
juvenile survival:
low
high
1
7
60
1000
60000
Migrants / generation (log scale)

(d)

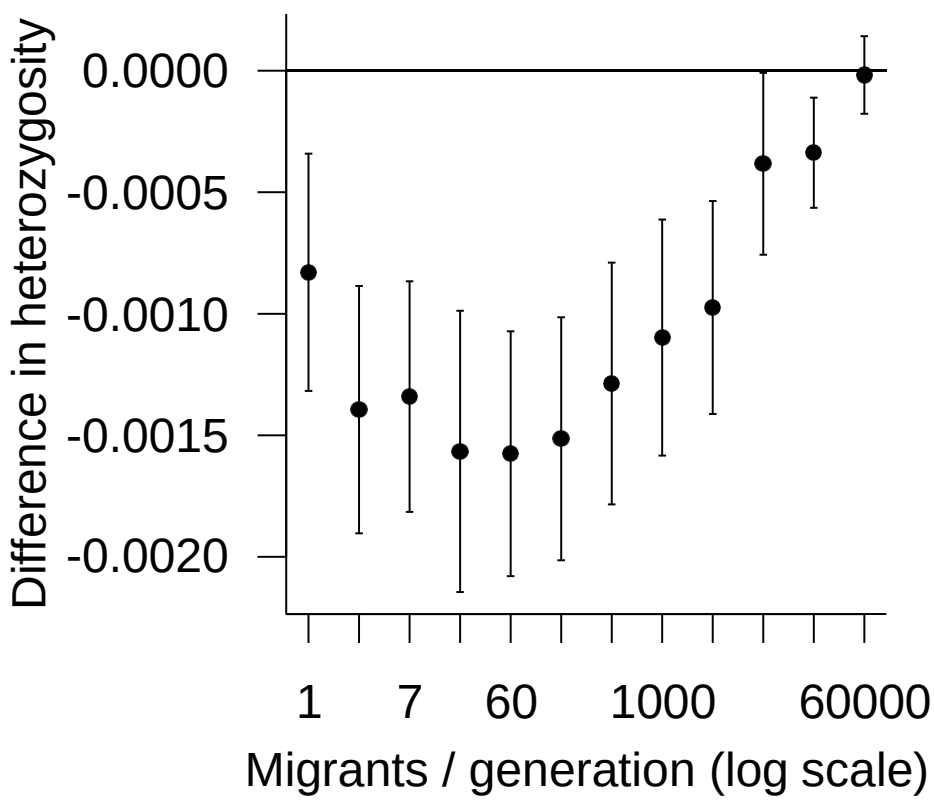

varying dispersal
Difference in heterozygosity
0.0000
-0.0005
-0.0010
-0.0015
-0.0020
1
7
60
1000
60000
Migrants / generation (log scale)

Figure 4. Genetic distance (F_{ST}) between pairs of *Plantago lanceolata* populations in the native European range was explained by two variables: (a) geographic distance and (b) distance in precipitation seasonality (coefficient of variation of annual mean precipitation) between sites. A generalised dissimilarity model indicated these variables had a significant (adjusted $P < 0.001$) effect on F_{ST}, given all other variables in the model (geographic distance, mean temperature, mean precipitation, temperature seasonality and precipitation seasonality). Deviance explained by the model was 74.3% and the model splines are shown in SI Appendix, Fig. S2.

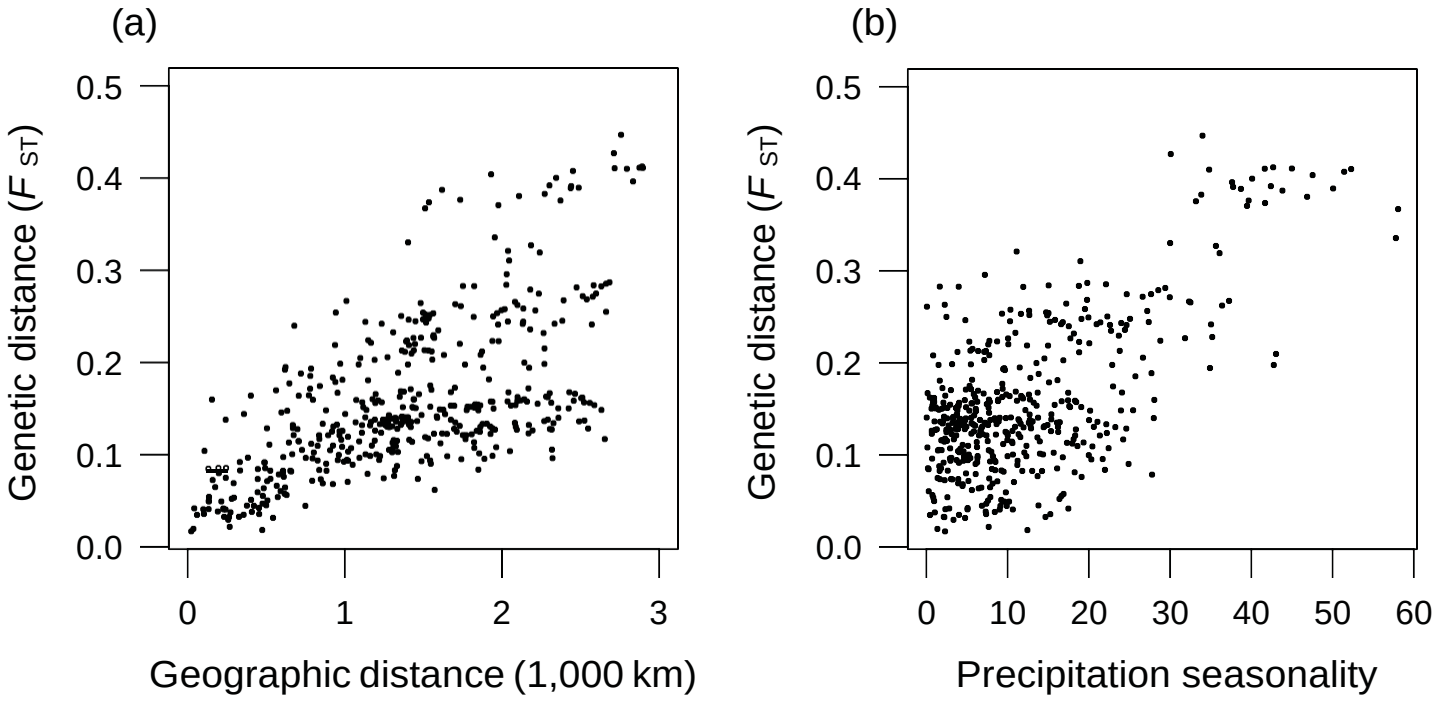

(a)
Genetic distance (F_{ST})
0.0
0.1
0.2
0.3
0.4
0.5
0
1
2
3
Geographic distance (1,000 km)
(b)
Genetic distance (F_{ST})
0.0
0.1
0.2
0.3
0.4
0.5
0
10
20
30
40
50
60
Precipitation seasonality

Figure 5. Environmental influences on demography and genetic diversity within populations in the native European (n = 30) and non-native (n = 14) range of *Plantago lanceolata* (model estimates and 95% confidence intervals shown over raw data). First-ranked models are shown for environmental influences on (a) population growth rate, (b) reproductive effort, (c–d) neutral genetic diversity and (e–f) adaptive genetic diversity. In all models except (e), the additive and interactive models both had support from the data (ΔAICc < 2, see SI Appendix, Fig. S3 and Dataset S1). For (e), the interaction between temperature seasonality (standard deviation of annual mean temperature at each site) and range (native/non-native) was the only model supported by the data (AICc weight = 0.95).

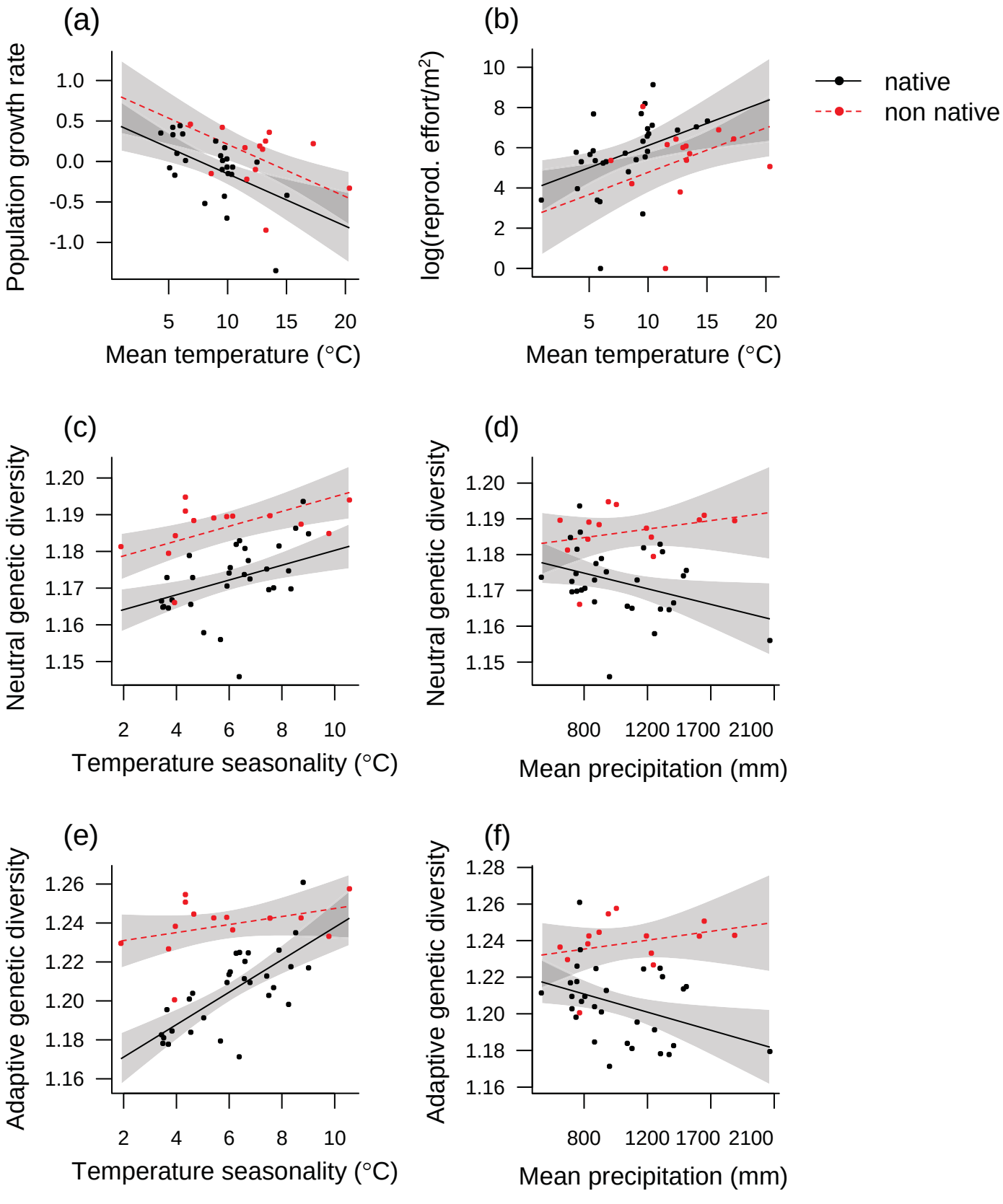

(a)
Population growth rate
1.0
0.5
0.0
-0.5
-1.0
5
10
15
20
Mean temperature (°C)
(b)
log(reprod. effort/m²)
10
8
6
4
2
0
5
10
15
20
Mean temperature (°C)
native
non native
(c)
Neutral genetic diversity
1.20
1.19
1.18
1.17
1.16
1.15
2
4
6
8
10
Temperature seasonality (°C)
(d)
Neutral genetic diversity
1.20
1.19
1.18
1.17
1.16
1.15
800
1200
1700
2100
Mean precipitation (mm)
(e)
Adaptive genetic diversity
1.26
1.24
1.22
1.20
1.18
1.16
2
4
6
8
10
Temperature seasonality (°C)
(f)
Adaptive genetic diversity
1.28
1.26
1.24
1.22
1.20
1.18
1.16
800
1200
1700
2100
Mean precipitation (mm)

Table 1. Demographic variables used to analyse population processes that are important to genetic diversity. The relevance of demographic variables to genetic diversity is outlined in Fig. 1 and described in detail by Ellegren and Gaultier[4]

Demographic variable measured	**Used as a proxy for**	**Relevance to genetic diversity**	**Formula**
Density	Population size	Effective population size	Number of rosettes / m^2 (N)
Reproductive effort per unit area	Fecundity	Fitness	(inflorescence length x no. flowering stems) / m^2
Empirical population growth rate	Combined effects of survival & fecundity	Fitness	$\log(N_{t+1}/N_t)$